e-4Cast

E-Commerce Analysis & Time-Series Demand Forecasting

By Janet TING

Key Questions ?

PROBLEM STATEMENT

Manage Inventory

Set Sales goals

Secure sales

Plan budget

Grow customers

PROJECT GOALS

Sales Orders Trend

Analyse datasets
Time series perspective

Forecasting Model

Test and build model to forecast sales demand for next 2 months

Forecast App

Forecast and visualize time-series data

METHODOLOGY

Data Preparation

Null values,
Duplicates,
Merging

EDA & Feature Engineering

Overview of trends
Time-series plots

Pre-Processing

Resampling to Day
Transforming Data (4)

Modeling

Stationary Test
Time-Series models (4)
Hyperparameter Tuning
Model Selection

Forecating model + Streamlit

Model Deploy

01

DATA PREPARATION

Olist E-Commerce Dataset

100K+ rows, 40 columns

Payment Records

Types, Method, Time

Orders Records

Timestamp of Purchase, Approval, Delivered, Received, Estimated

Orders_Items

IDs, Time

Sellers Records +

IDs, Geolocation

Customers Records +

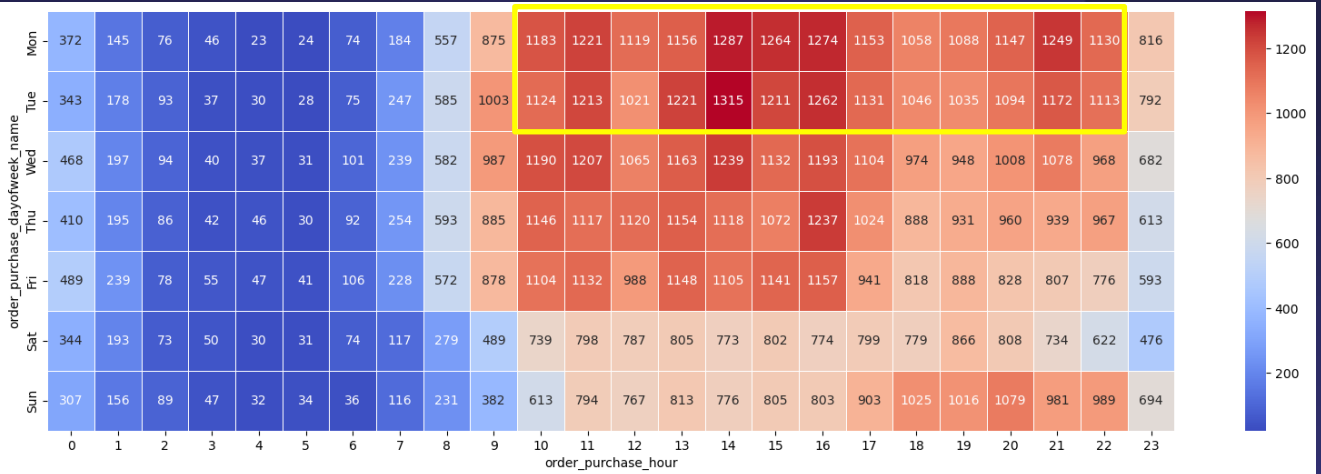
IDs, Geolocation

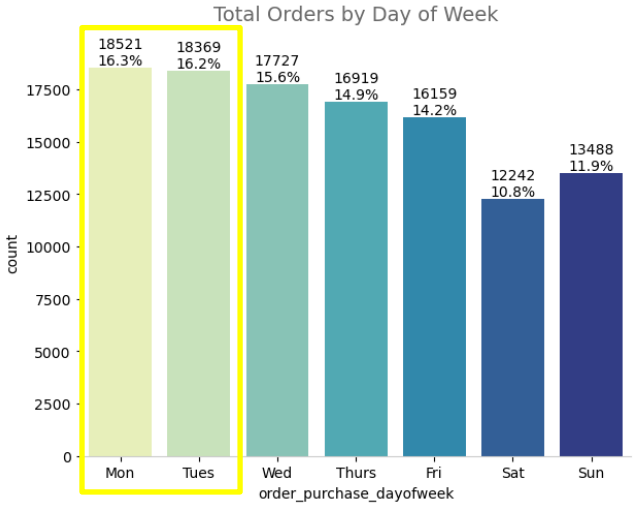
Products Records

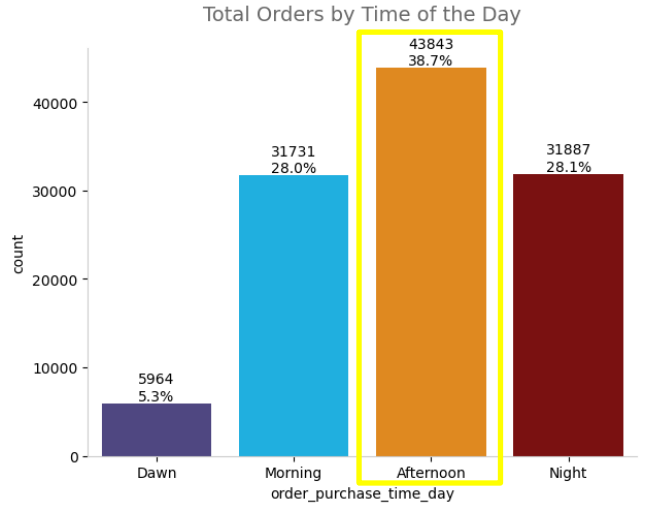
Product categories, IDs, Descriptions, Size

02 EDA & FEATURE ENGINEERING

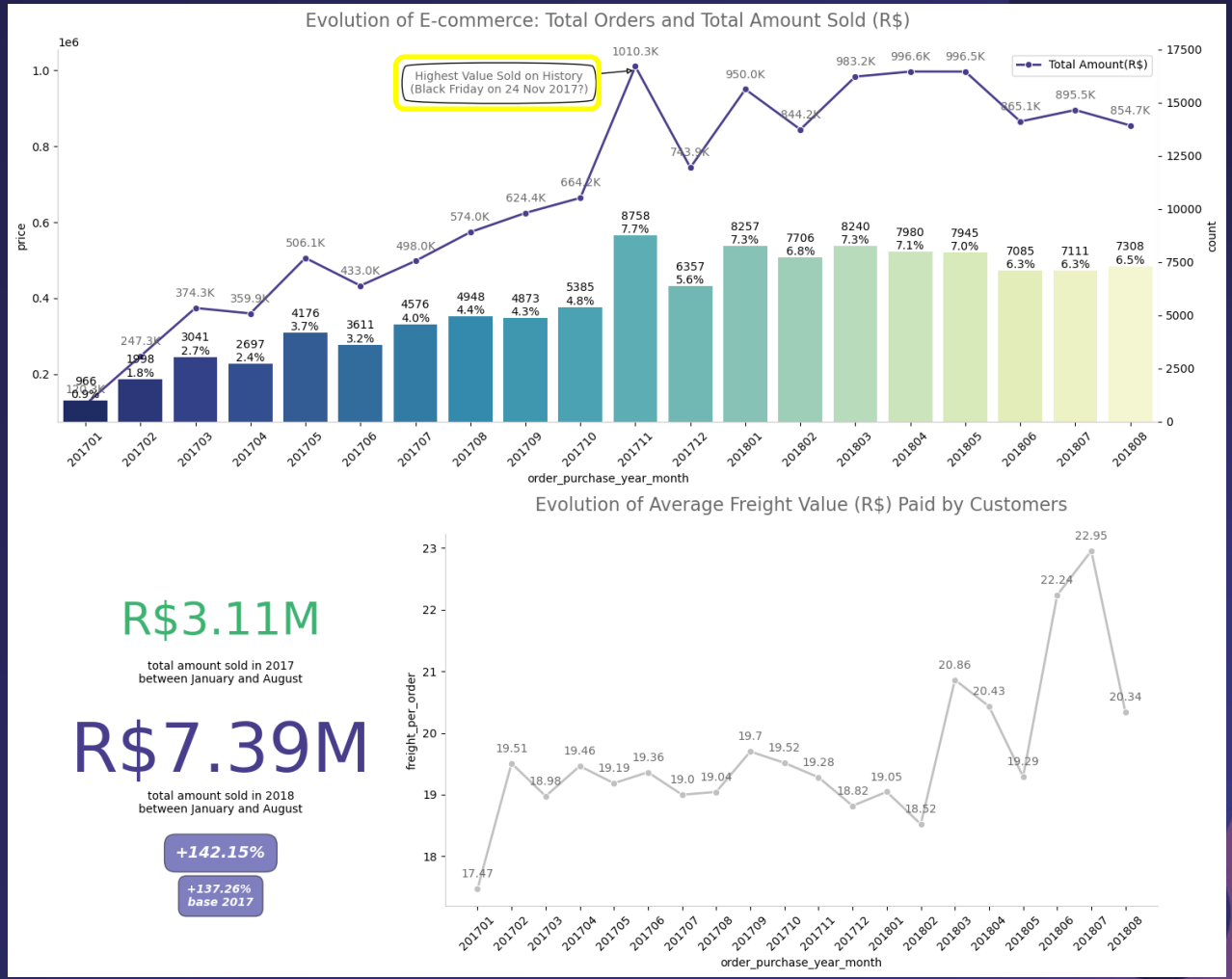
E-Commerce Purchase Order Statistic







E-Commerce Sales Trends in Brazil

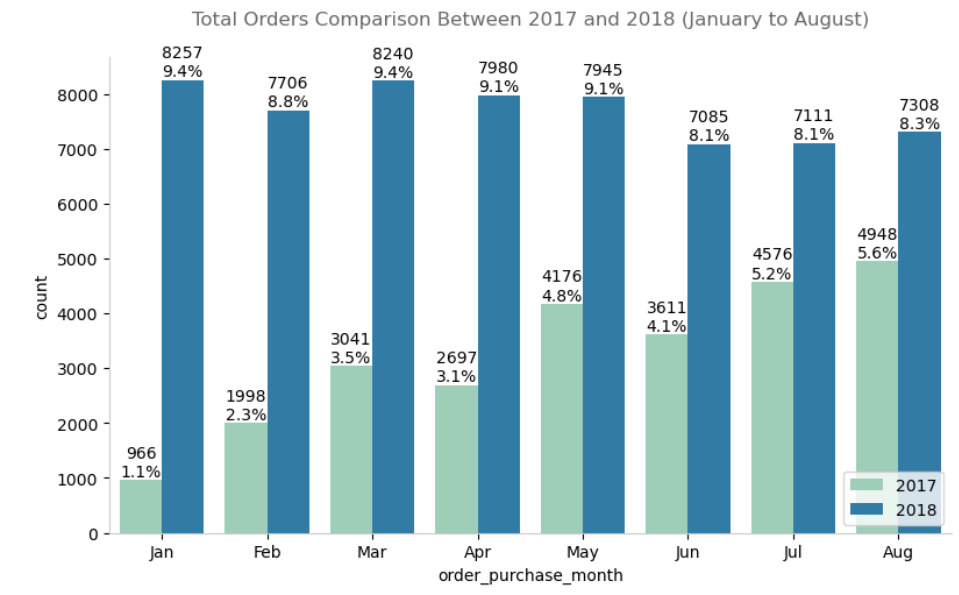


Comparison between 2017 and 2018

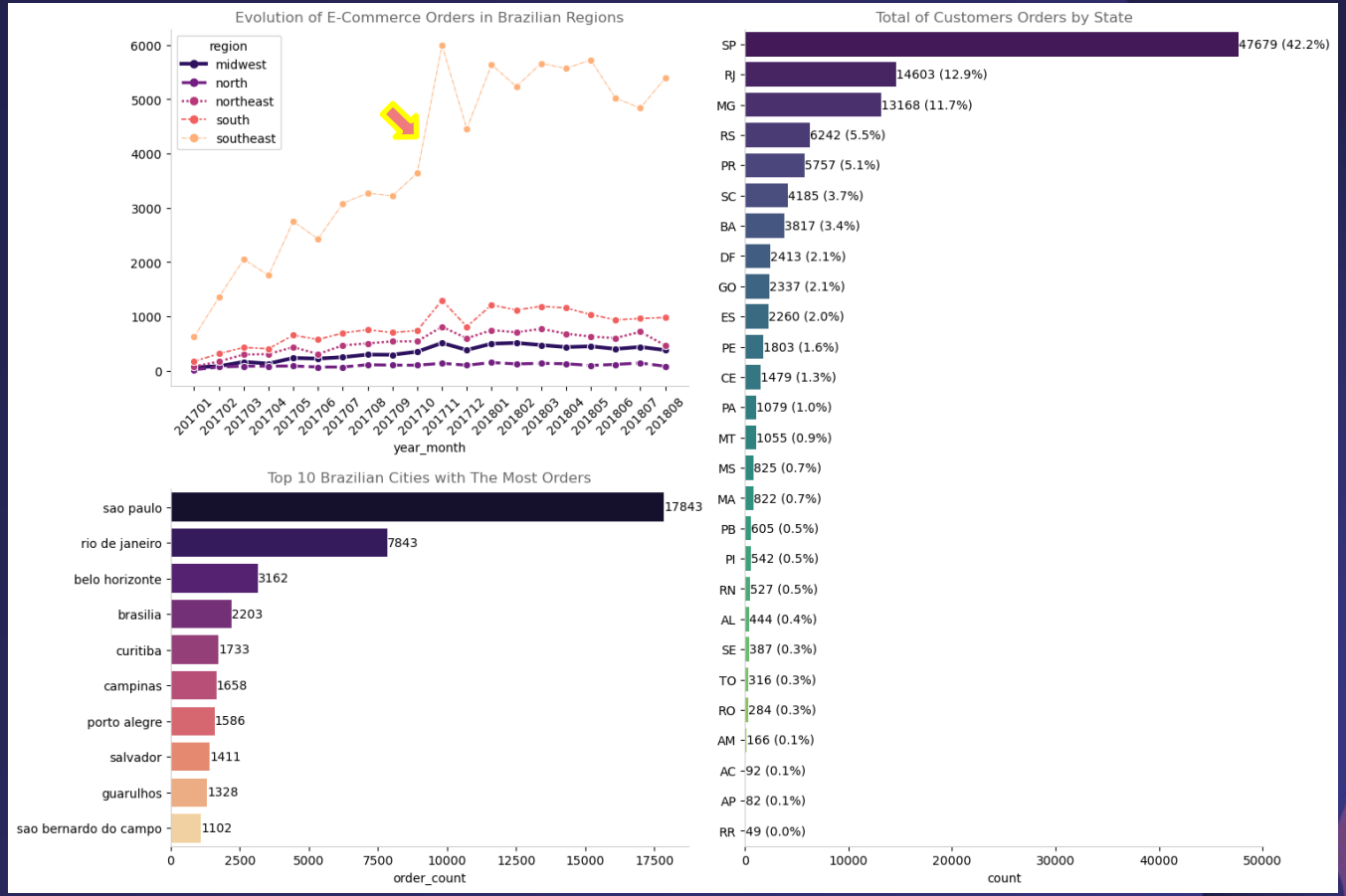
26013
orders registered in 2017
between January and August

61632
orders registered in 2018
between January and August

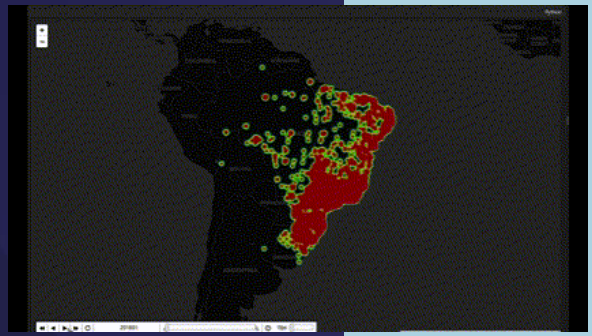
+142%



E-commerce Evolution in Brazil by Location

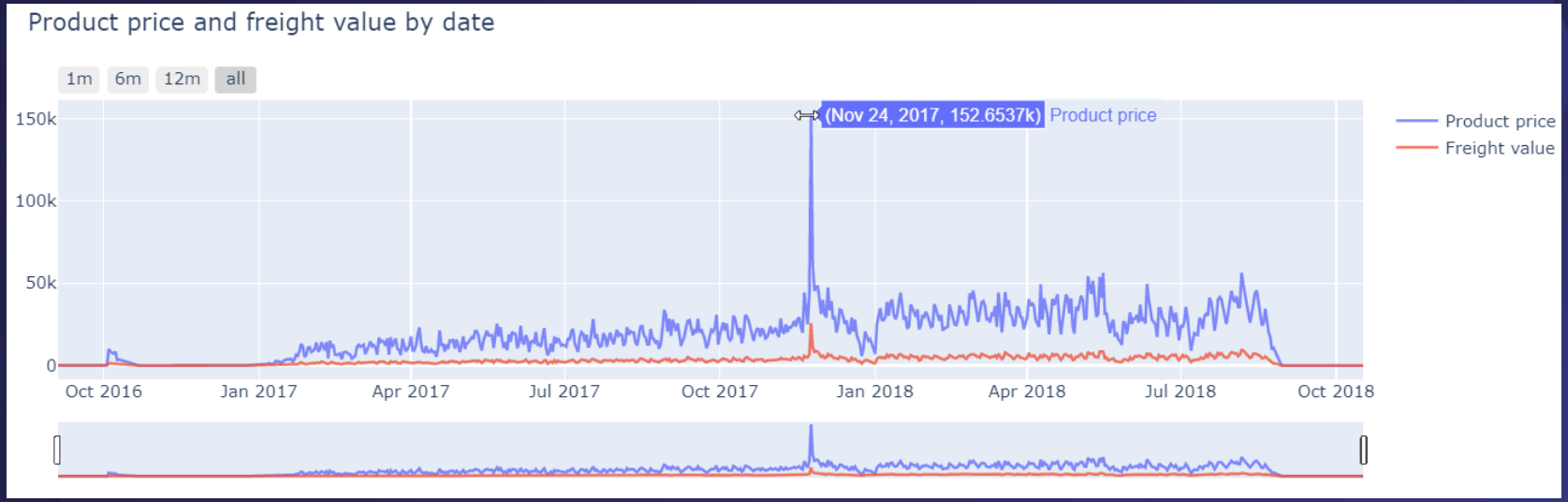


Customers Distribution in Brazil



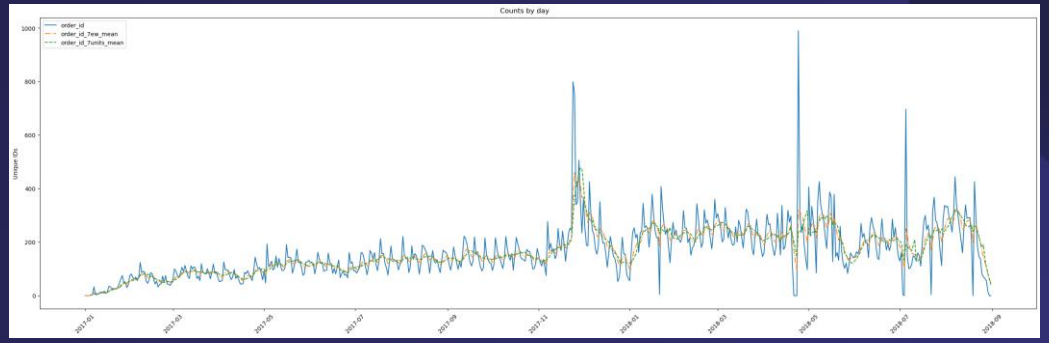
Timeframe: Jan-Aug 2018
*red region – high concentration

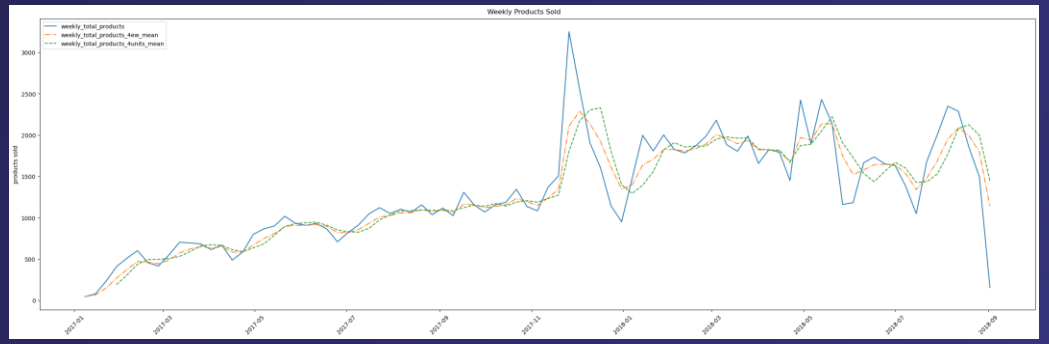
Product price and freight over time

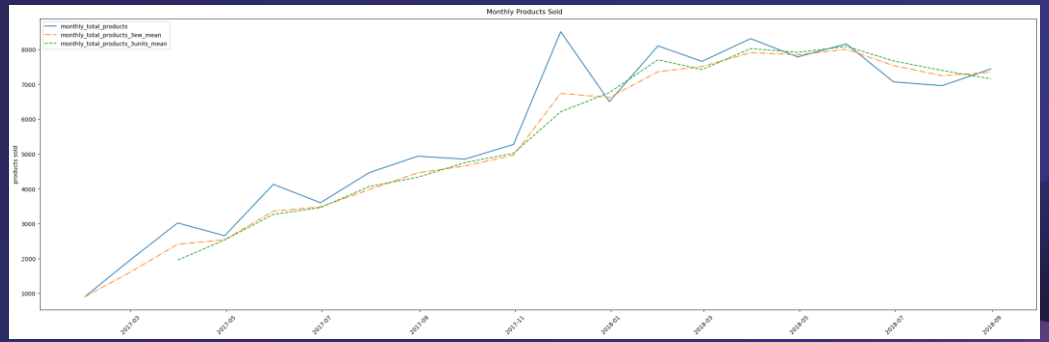


Obvious outliers:
24 Nov 2017 – Black Friday

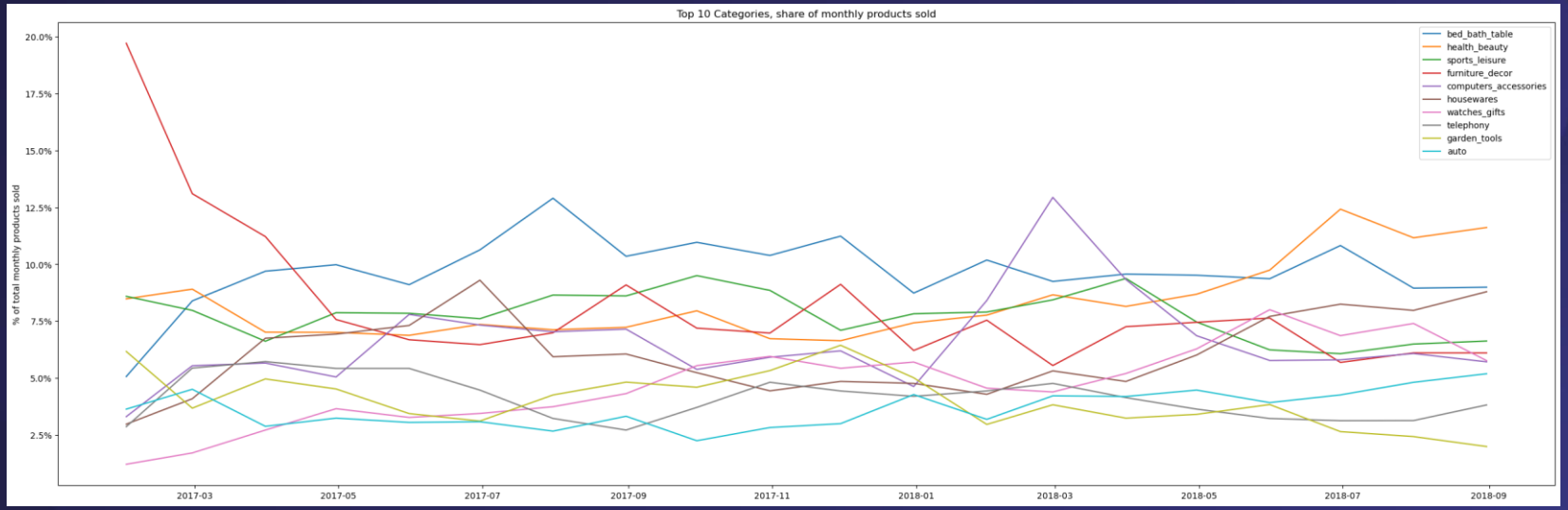
Daily, Weekly, and Monthly Products Sold with Rolling Means







Top 10 Product Categories Share of Monthly Products Sold



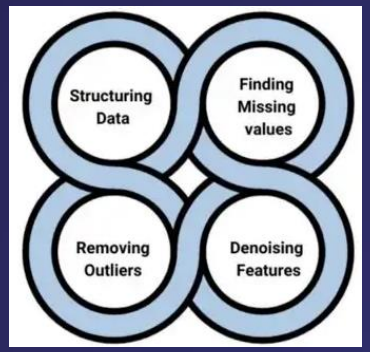
*Furniture_decor's monthly share drops early 2017
*Computer_accessories' monthly share spikes in March 2018

03

PRE-PROCESSING & MODELLING

Time Series Model Preprocessing

	A	B	C
1		order_purchase_timestamp	order_id
2	0	2/10/2017 10:56	e481f51cbdc54
3	1	24/7/2018 20:41	53cdb2fc8bc7d
4	2	8/8/2018 8:38	47770eb9100c2
5	3	18/11/2017 19:28	949d5b44dbf5d
6	4	13/2/2018 21:18	ad21c59c0840e
7	5	9/7/2017 21:57	a4591c265e18c
8	6	11/4/2017 12:22	136cce7faa42f



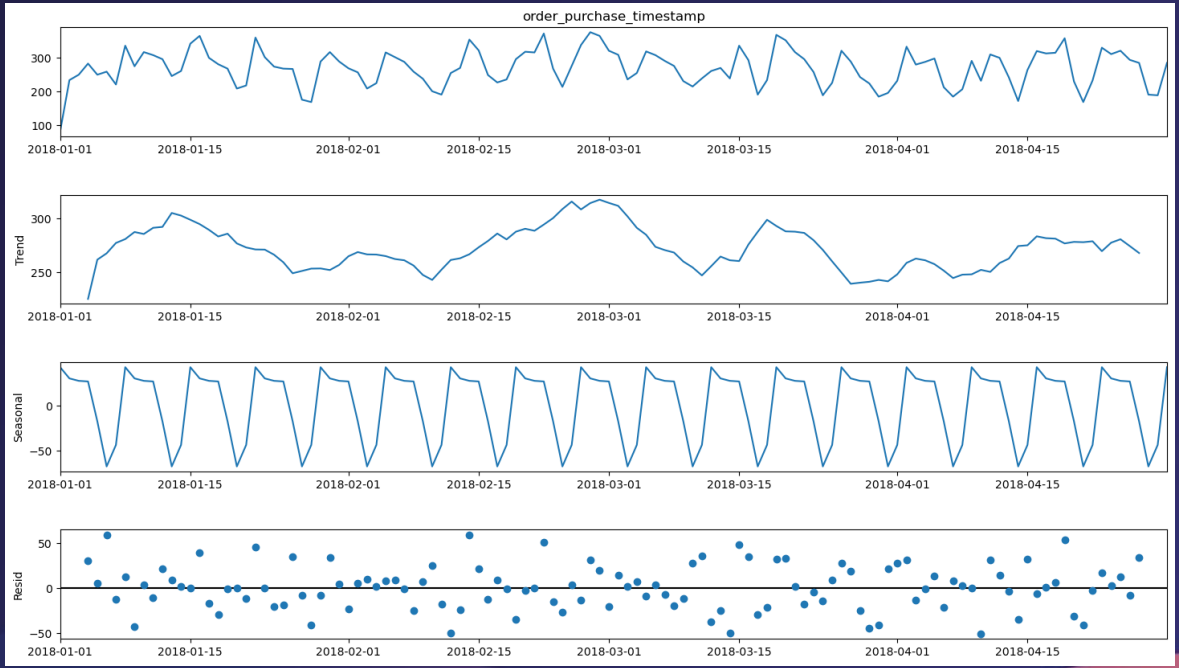
	A	B
1	ds	y
2	1/1/2017	0
3	2/1/2017	0
4	3/1/2017	0
5	4/1/2017	0
6	5/1/2017	32
7	6/1/2017	4
8	7/1/2017	5

Original Data

Sequence does matter!

Processed data

Seasonal Decomposition



Observations:

- Inconsistent trends
- Strong weekly seasonality
- Noisy Residual

PRE-PROCESSING & MODELLING

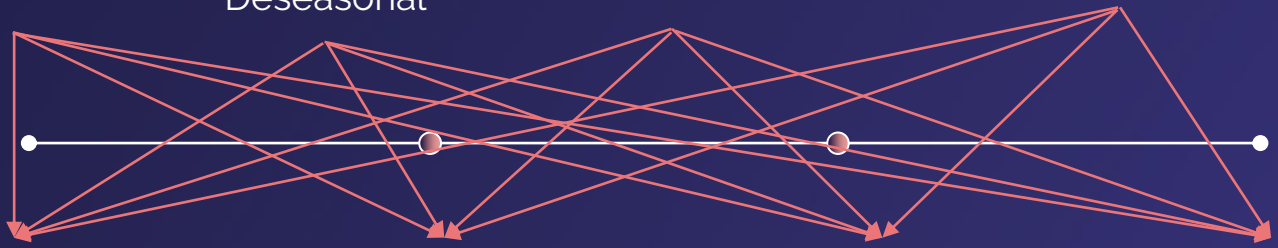
Tested 16 models
*5-fold cross validation

Original
-

Seasonal Trend based on Loess (STL)
Deseasonal

Box-Cox
Residual

Seasonal Differencing
-



Time Series Regression
-

Holt Winter's Exponential Smoothing (HWES)
-

Auto (S)ARIMA
-

Long-short term Memory (LSTM)
-

Performance Metrics

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

- Average distance from the predicted value (magnitude of error)
- The lower RMSE, the better model performance.

- Scale dependent of Y
- Heavily affected by outliers

Symmetric Mean Absolute Percentage error (SMAPE)

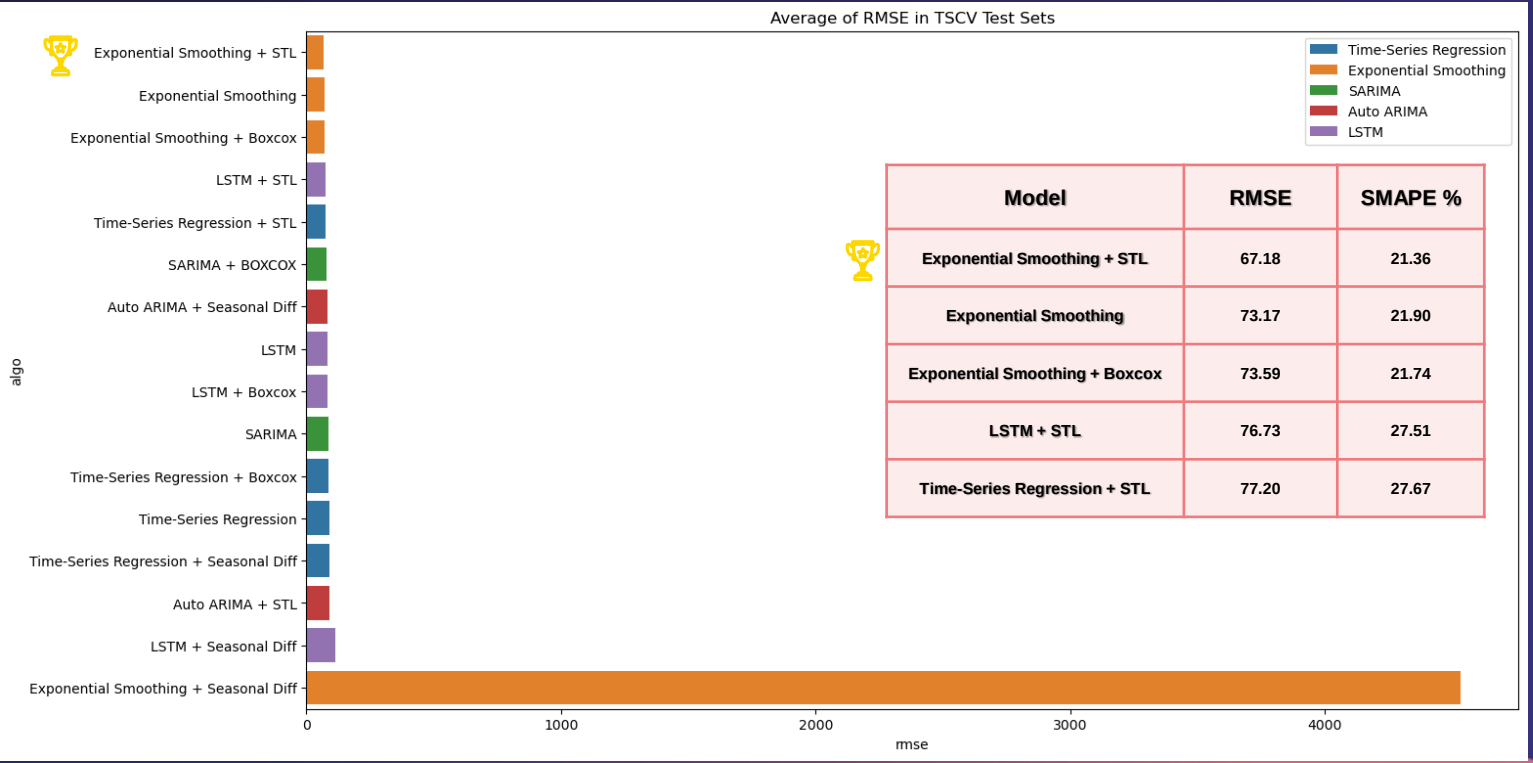
$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

where A_t is the actual value and F_t is the forecast value.

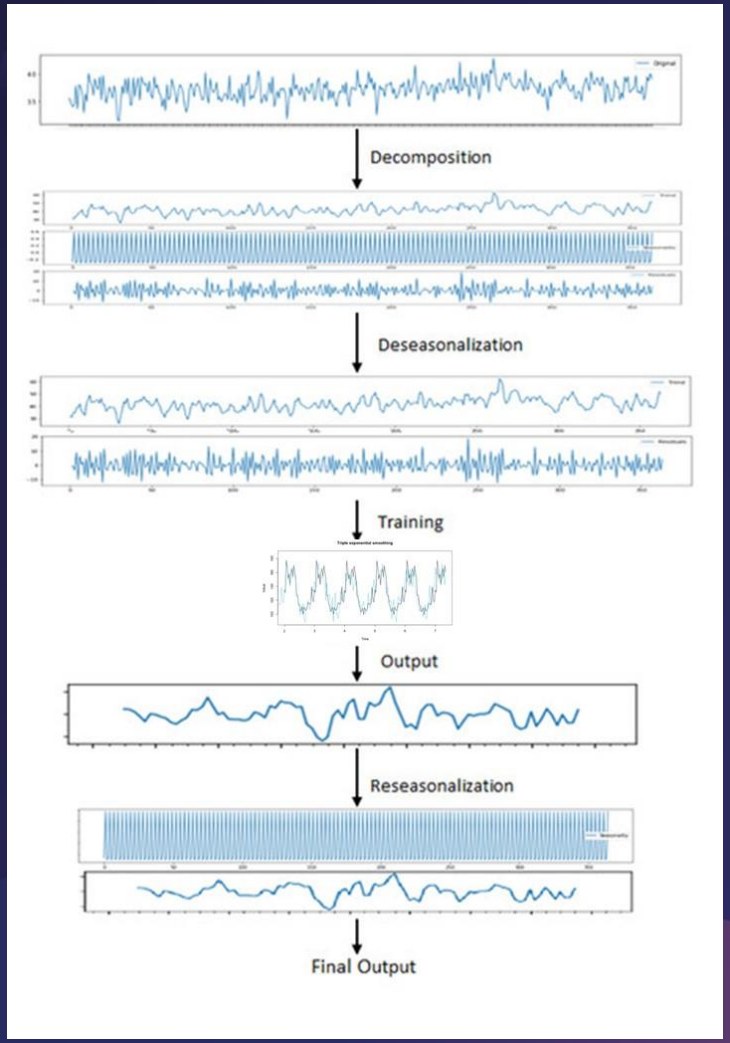
- Fixed the shortcoming of original MAPE when value is as small as 0
- The lower %, the better model performance

- If underpredict, the % is higher

EVALUATION



Model	RMSE	SMAPE %
Exponential Smoothing + STL	67.18	21.36
Exponential Smoothing	73.17	21.90
Exponential Smoothing + Boxcox	73.59	21.74
LSTM + STL	76.73	27.51
Time-Series Regression + STL	77.20	27.67



MODEL UNDERSTANDING – Exponential Smoothing + STL

Data: STL transformed data (remove seasonalities).

Model: Holt-Winters/ Triple Exponential Smoothing (for univariate data)

- explicitly uses weight averages of past observations.
- weights decaying exponentially as the observations get older.

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

Future Forecast

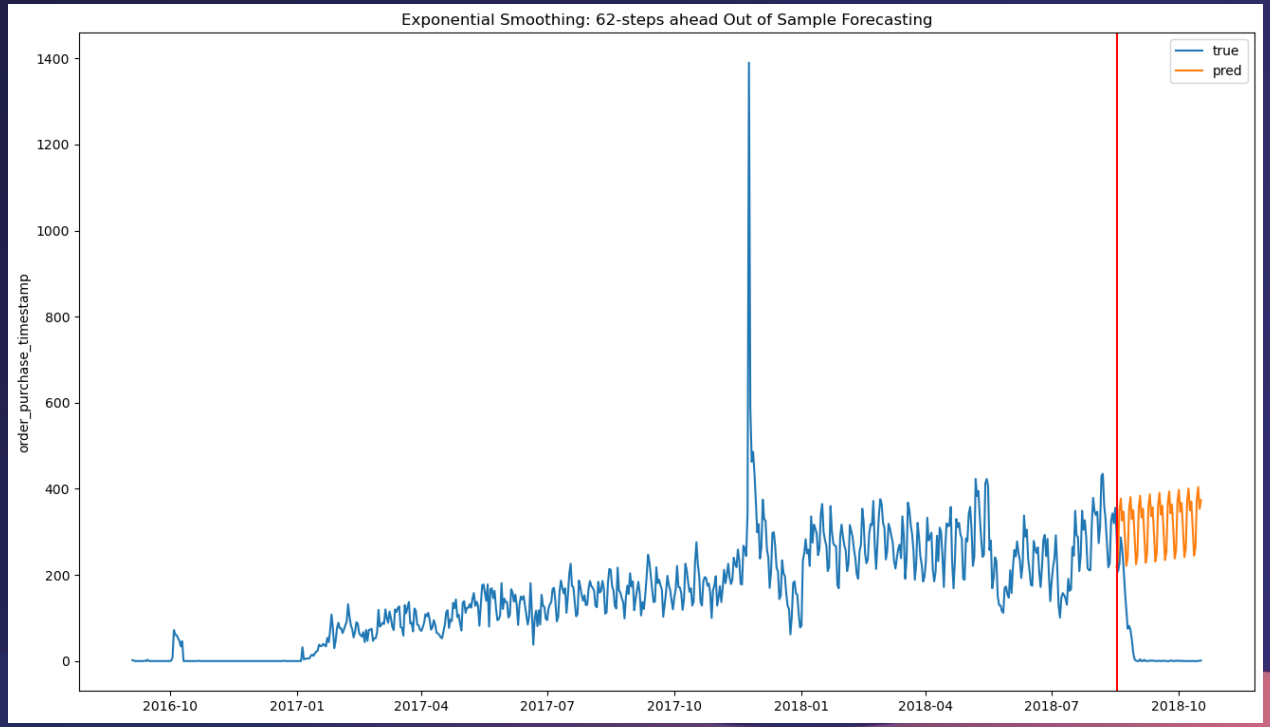
Level Smoothing(α)

Trend smoothing (β)

Seasonal smoothing(γ)

Pipeline: Grid search over the parameters (smallest RMSE).

2-Month FORECAST



Forecast upward trend rather than original data

News about Olist being backed by investment and its expansion.

- source: https://www.bloomberg.com/news/articles/2020-12-21/softbank-backed-olist-buys-brazil-logistics-firm-in-online-push
- https://www.reuters.com/technology/goldman-sachs-redpoint-finance-new-round-brazils-olist-2021-04-15/

LIMITATIONS

1) **The selected models are Univariate only.**
2) **No exogenous variables added (holidays, product nature, geolocation, etc).**
3) **Yet to test on other advanced ts-forecasting models**
4) **Limited data (2017-01-01 to 2018-08-17).**

RECOMMENDATION

1) Businesses manage more aspects of the supply chain, reduces 10% lower weekly error
2) Or consider expanding vendor-managed inventory programs and leveraging such data
3) Practice more proactive demand forecasting (i.e. both historical data & external factors)
4) Storesellers examine product category/ies on a national level, taking in economic factors (e.g. employment and cost of living)

LIMITATIONS

1) The selected models are Univariate only.
2) No exogenous variables added (holidays, product nature, geolocation, etc).
3) Yet to test on other advanced ts-forecasting models
4) The current complete data available is only from 2017-01-01 to 2018-08-17.

RECOMMENDATION

1) **Businesses increase aspects of the supply chain, reduces 10% lower weekly error*.**
2) **Expanding vendor-managed inventory programs and leveraging data**
3) **Practice more proactive demand forecasting (both historical data & external factors.)**
4) **Store-sellers examine product categories on a national level, taking in economic factors (employment and cost of living)**

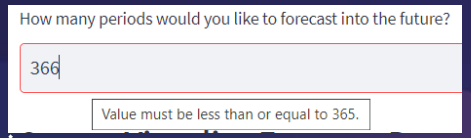
DEPLOYMENT

User Input

Upload a tablular file in correct format (contains time & target column):

	A	B
1	ds	y
2	1/1/2017	0
3	2/1/2017	0
4	3/1/2017	0

Select forecast horizon (0-365)



Process

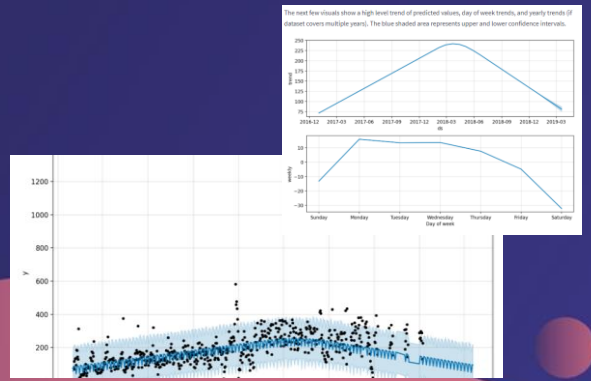
Streamlit processes and analyzes data seamlessly.

It loads and fits the model with data given.

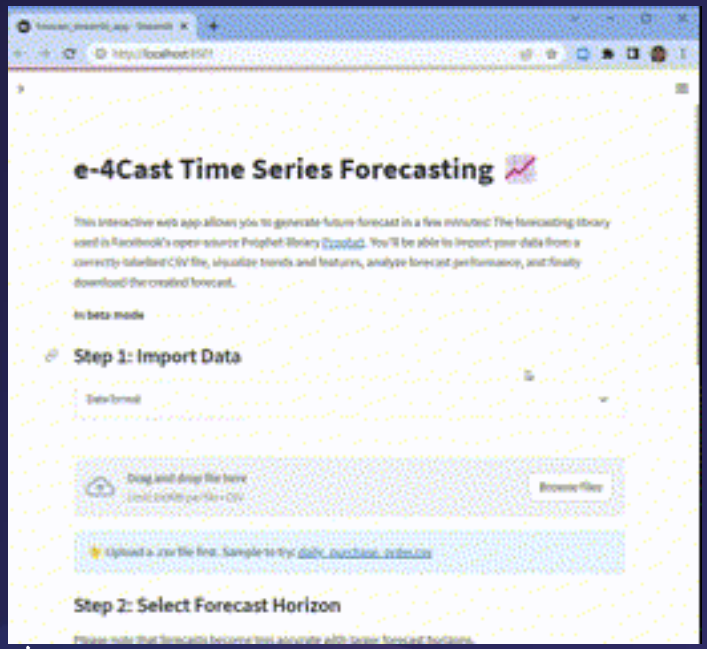
Output

Visualization through plots and tables are shown.

Forecast were generated and can be downloaded.



Streamlit Demo



Steps:

1. Import Data
2. Select Forecast Horizon
3. Visualize Forecast Data
4. Download the Forecast

CONCLUSIONS

- Enables sales team to set realistic short-term goal which leads to securing of sales.
- Enables data-driven inventory management.
- Enables budget planning optimization
- Enables better customer segmentation profiling.

THANKS!

Do you have any questions?

https://github.com/Janet-THY

janettinghy@gmail..com